PLISKA
STUDIA MATHEMATICA BULGARICA

ПЛИСКА
БЪЛГАРСКИ МАТЕМАТИЧЕСКИ СТУДИИ

Pliska Stud. Math. Bulgar. **12** (1998), 227-234

PLISKA
STUDIA MATHEMATICA
BULGARICA

PRINCIPLES OF EXTREMUM AND APPLICATION TO SOME PROBLEMS OF ANALYSIS

V. M. Tikhomirov

The aim of this paper is to demonstrate applications of a direct approach to the solution of extremal problems to some concrete problems of classical analysis, calculus of variations and approximation theory.

Keywords: Lagrange principle, duality, inequalities for derivatives.

AMS subject classification: 41A17, 41A50, 49Kxx, 90C25

1 On some principles of the theory of extrema

We will use the following procedure of investigation of extremal problems. At first we *formalize* the problem, i.e. we express the problem by means of equalities, inequalities and inclusions.

Then we *use Lagrange principle for formulating necessary conditions of extremum* or *formulate the dual problem* (if the initial problem was convex).

After that we *investigate the equations*, which were obtained after application of the Lagrange principle or *solve the dual problem.*

And at last we formulate the final result.

Let us give a more precise explanation of "Lagrange principle" and "dual problem".

The first method of solution of extremal problems goes back to Fermat. Here it is: *if* $f(x) \to \mathrm{extr}$ *is a smooth (or convex) problem without constraints and* $\widehat{x}$ *is a solution of the problem, then*

$$f'(\widehat{x}) = 0 \quad (0 \in \partial f(\widehat{x})).$$

($\partial f(\widehat{x})$ is the subdifferential of f at $\widehat{x}$.)

Since 1759, Lagrange began to investigate extremal problems with constraints. Lagrange formulated the following general idea for finding extremum of an extremal problem. His idea can be expressed in the following form: *if one looks for the maximum or minimum in a problem with constrains, it is necessary to form the Lagrange function and after that write the necessary condition "as if the variables are indepndent".*

We modify the main idea of Lagrange for the following extremal problem:

$$(P), \qquad f_0(x,u) \to \min, \quad f_i(x,u) \le 0, \quad 1 \le i \le m, \quad F(x,u) = 0,\ u \in U$$

where $f_i : X \times U \to \mathbb{R} \cup \{+\infty\}$, X (usually) is a normed space, U is some set, $F : X \times U \to Y$, where Y also is a normed space.

The function $\mathcal{L}(x) = \sum_{i=0}^{m} \lambda_i f_i(x) + \langle y^*, F(x,u)\rangle$ is called *the Lagrange function* of this problem, the numbers λ_i are called *Lagrange multipliers.*

Let functionals and mappings be smooth over variables of the first group and be convex over the second one and let $(\widehat{x}, \widehat{u})$ be a solution of the problem, then (according to Lagrange's idea) "it suffices" to write the necessary condition in the (smooth) problem

$$\mathcal{L}(x, \widehat{u}) \to \min$$

"as if the variables are independent" (in other words to apply the Fermat theorem) and besides that we must apply a criterium for the solution of the convex problem

$$\mathcal{L}(\widehat{x}, u) \to \min, \quad u \in U$$

(together with *conditions of complementary slackness*: $\lambda_i \ge 0$, $1 \le i \le m$).

We call this procedure *Lagrange principle.* It is important to remark, that in convex case necessary conditions coincide with sufficient one, i. e. Lagrange principle has the most complete form: *the solution of the problem is an absolute minimum of the Lagrange function.*

Along with Lagrange principle, we use some results of convex analisis.

The most important of them is one of the form of duality in the convex programming, namely *the principle of constructing of the dual problem.*

If a prime problem is: $f(x) \to \min$ $(f : X \to \mathbb{R} \cup \{+\infty\})$ and the function $F : X \times Y \to \mathbb{R} \cup \{+\infty\}$, $f(x) = F(x, 0)$ is the *perturbation* of f, then the dual problem has the form $g(y^*) \to \max, g(y^*) = -F^*(y^*, 0)$ where F^* is a Legendre transform of F.

Besides all this, we will use the following two important theorems.

Theorem of Dubovitski-Milyutin. *Let f_1 f_2 be convex and continuous functions at x and $f_1(x) = f_2(x)$. Then $\partial(f_1 \vee f_2)(x) = \mathrm{co}\{\partial f_1(x) \cup \partial f_2(x)\}$ (*co *is the convex hull, $(f_1 \vee f_2)(x) = \max(f_1(x), f_2(x))$.*

Decomposition theorem. *Let T be a compact, $F : T \times \mathbb{R}^n \to \mathbb{R}$, $F = F(t,x)$. Let F be upper semicontinuous over t for all x and convex over x for all t. Then there exists a number $r \le n+1$ and r points $\{\tau_i\}_{i=1}^r$, $\tau_i \in T$, such that*

$$\inf_x \sup_t F(t,x) = \inf_x \max_{1 \le i \le r} F(\tau_i, x)$$

See about all this [1].

2 Solution of concrete problems

2.1 Tchebyshev alternance theorem and its generalizations

In the first article devoted to approximation theory Tchebyshev formulated a necessary condition for an algebraic polynomial of the best approximation of a given continuous function (in uniform metric). We solve this problem using considerations of p.1.

1. Formalization.

$$(i) \qquad f(x) := \max_{t\in[t_0,t_1]}\left|x(t)-\sum_{k=1}^{n+1}x_kt^{k-1}\right| \to \min, \quad x=(x_1,\cdots,x_{n+1}).$$

This is a convex problem without constraints. Existence of a solution $\widehat{x} \Leftrightarrow \widehat{p}(\cdot) = \sum\limits_{k=1}^{n+1}\widehat{x}_kt^{k-1}$ follows from the principle of compactness.

2. Fermat's theorem leads to the inclusion $0 \in \partial f(\widehat{x})$.

3. Investigation. From the decomposition theorem it follows that there exists a natural number $r \le n+1$ and r points $\{\tau_i\}_{i=1}^r$ on the segment $[t_0,t_1]$ such that

$$(ii) \qquad f(\widehat{x}) = |y(\tau_i)|,\ 1 \le i \le r, \quad y(t) = \left|x(t)-\sum_{k=1}^{n+1}\widehat{x}_kt^{k-1}\right|.$$

We see that r functions $f_i(x) = |x(\tau_i) - \sum\limits_{k=1}^{n+1} x_k\tau_k^{k-1}|$ attain the unique value $f(\widehat{x})$ at the point $\widehat{x}$, and this point is a solution of the problem (i). From Dubovitsky-Milyutin theorem it follows that zero vector (which belongs to $\partial f(\widehat{x})$) is represented as a convex hull of $\partial f_i(\widehat{x}) = f'(\widehat{x}) = \mathrm{sgn} y(\tau_i)(1,\tau_i,\ldots,\tau_i^n)$. Hence

$$(iii) \qquad \sum_{i=1}^{r}\alpha_i\mathrm{sgn}(x(\tau_i)-\sum_{k=1}^{n+1}\widehat{x}_k\tau_i^{k-1})(1,\tau_i,\cdots,\tau_i^n) = 0, \quad \alpha_i>0, \quad \sum_{k=1}^{r}\alpha_i = 1$$

We see that the homogeneous system with $n+1$ equations and $r \le n+2$ unknowns (and determinants of Wandermond type) has nonlinear solution. Hence $r = n+2$, and from the explicit expression of solution of these equations it is easy to show that numbers $\alpha_i\mathrm{sgn}(x(\tau_i)-\widehat{p}(\tau_i))$ change their signs.

We have proved

Tchebyshev alternance theorem. *The polynomial $\widehat{p}(\cdot) = \sum\limits_{k=1}^{n+1}\widehat{x}_kt^{k-1}$ is a polynomial of the best approximation of a function $x(\cdot)$ in $C([t_0,t_1])$ iff there exist $n+2$ points in which the function $x(\cdot)-\widehat{p}(\cdot)$ obtains its maximum and minimum values with alternation.*

Remark. The analogous considerations immediately lead to criterium of an element of the best approximation of a function from $C(T,Y)$, (where T is a compact and Y is a normed space or even Y may be a linear space with Minkovsky metric) by an arbitrary convex set. This result generalizes many previous theorems of Bernstein, Kolmogorov and others (see [2]).

2.2 Tchebyshev extrapolation problem

The statement of the problem is the following. Let a polynomial of degree n be bound by a definite constant on a segment of the real line. The question is: *what are the limits of a value of the polynomial in a fixed point?*

1. Formalization

$$(i) \qquad x(\tau) \to \max, \quad \max_{t\in[-1,1]} |x(t)| \le 1, \quad x(t) = \sum_{k=1}^{n+1} x_k t^{k-1},$$

$$|\tau| > 1, \quad x(\cdot) \Leftrightarrow x = (x_1, \dots, x_{n+1}).$$

It is a problem of convex programming. From the principle of compactness it follows that a solution $\widehat{x}$ of the problem exists.

2. Lagrange principle here (because of convexity of the problem) has the following form: *Lagrange function*

$$\mathcal{L} = f(x) + \lambda g(x), \quad f(x) = \sum_{k=1}^{n+1} x_r \tau^{k-1}, \quad g(x) = \max_{t\in[-1,1]} \left| \sum_{k=1}^{n+1} x_k t^{k-1} \right|$$

attains its absolute minimum at $\widehat{x}$.

3. Investigation. Applying the decompozition and Dubovitsky-Milyutin theorems, we come to the following identity

$$-x(\tau) + \lambda \sum_{k=1}^{r} \alpha_r \operatorname{sgn}\widehat{x}(\tau_k) x(\tau_k) = 0 \quad \forall x(\cdot) \in \mathcal{P}_n,$$

$$r \le n+2, \quad \alpha_k > 0, \quad 1 \le k \le r, \quad \sum \alpha_k = 1,$$

$$-1 \le \tau_1 < \cdots < \tau_r \le 1, \quad |\widehat{x}(\tau_k)| = 1.$$

It is evident that $r \neq n+2$ (otherwise, the polynomial $\dot{\widehat{x}}(\cdot)$ of degree $n-1$ has n zeroes at $\{\tau_k\}_{k=2}^{n+1}$, but it is impossible).

If we assume that $r < n+1$, then we come to contradiction substituting the polynomial $x_1(t) = \prod_{k=1}^{r} (t - \tau_k)$ into our identity. Hence $r = n+1$ and consequently the solution is the polynomial $\cos n \arccos t$. We have proved

Theorem on extrapolaton of polynomials. *Tchebyshev polynomial $T_n(\cdot)$ gives the solution of the problem of extrapolation.*

Remarks

1. The similar results hold for an arbitrary T-system in the space $C([t_0, t_1])$ with a weight: *the solution of extrapolation problem is in some sence analog of Tchebyshev polynomials.*

2. If in the problem (i) τ is a complex number (but coefficients are real) then the solution of the problen is either Tchebyshev or Zolotarev polinomials. These and some other questions are discussed in the paper [3].

2.3 A. Markov's problem on polynomial derivatives on a fixed point.

The initial information about a polynomial is the same as in p.2.2. The problem is: *what are the limits for a value of a derivative of the polynomial at a fixed point.*

1. Formalization.

$$f(x) = \dot{x}(\tau) \to \max, \quad f_1(x) = \max_{t \in [-1,1]} |x(t)| \le 1,$$

$$x(t) = \sum_{k=1}^{n+1} x_k t^{k-1}, x(\cdot) \Leftrightarrow x = (x_1, \ldots, x_{n+1}).$$

This is also the problem of convex programming and the solution of it exists.

2. Lagrange principle leads to the identity

$$(i) \qquad \dot{x}(\tau) + \lambda \sum_{j=1}^{r} \alpha_j \mathrm{sgn}\widehat{x}(\tau_j) x(\tau_j) = 0,$$

where $r \le n+2, \quad \alpha_j \ge 0, \sum_{j=1}^{r} \alpha_j = 1$ and $|\widehat{x}(\tau_j)| = 1$.

3. Investigation.

Similarly to the previous case, it can be proved that $r \ne n+2$ and $r \ge n$.

Let us show that extremal polynomial has n-alternance. If $\widehat{x}(\tau_j) = \widehat{x}(\tau_{j+1})$, we substitute the polynomial $x_2(t) = (t-\tau)^2 \prod_{k \ne j, j+1} (t - \tau_k)$ into (i) and come to contradiction.

(Polynomials of degree n which have n-alternance were described by Zolotarev. They are called *Zolotarev polynomials.*)

We have proved

Theorem of A. Markov on polynomial derivatives at a fixed point. *Solution of the problem coincides either with Tchebyshev or Zolotarev polynomial.*

Remark. It is very easy to describe *all* functionals on an arbitrary T-space for which the extremal polynomial is (generalised) Tchebyshev polynomial. In the partiqular case of A. Markov's problem, this criterion is equivalent to well known Markov's criterion.

In [3] we discuss some analogs of Markov's problem for ECT-subspaces of $C([t_0, t_1])$. (Nonzero function on T-space degree n have not more than $n-1$ zeroes; ECT is generalized Markov system.)

2.4 Inequalities for derivatives on line and half line.

The goal here is to find the best constant in the inequality

$$(*) \qquad \|x^{(k)}(\cdot)\|_{L_q(T)} \leq K\|x(\cdot)\|^{\alpha}_{L_p(T)}\|x^{(n)}(\cdot)\|^{\beta}_{L_r(T)}$$

(where $0 \leq k < n \quad n \in \mathbb{N}$, $1 \leq p,q,r \leq \infty$, $\alpha,\beta \geq 0$, $T = \mathbb{R}$ or $\mathbb{R}_+$), $x(\cdot) \in L_p(T)$, $x^{(n-1)}(\cdot)$ is locally absolutely continuous on T and $x^{(n)}(\cdot) \in L_r(T)$.

When T is fixed, the inequality $(*)$ depends on 5 parameters: n, k, p, q and r (α and β are equal to $\alpha = (n-k-1/r+1/q)/(n-1/r+1/p)$, $\beta = 1-\alpha$). The best constant in $(*)$ we denote by $K_T(n,k,p,q,r)$.

First results: $T = \mathbb{R}_+$, $n = 2$, $k = 1$, $p = q = r = \infty$ and $T = \mathbb{R}$, $n = 2$, $k = 1$, $p = q = r = \infty$ are due to E. Landau (1913) and Hadamard (1914). One of the most interesting results ($T = \mathbb{R}$, $p = q = r = \infty$, i.e for all $n \geq 2$ and $0 < k < n$) was considered and solved by Kolmogorov.

Analogous general inequalities were obtained only in six cases:

1) $p = q = r = 2, T = \mathbb{R}$ — Hardy-Littlewood-Polya,
2) $p = r = 2$, $q = \infty, T = \mathbb{R}$ — Taykov,
3) $p = q = 2, q = \infty, T = \mathbb{R}_+$ — Gabushin,
4) $p = q = r = 2, T = \mathbb{R}_+$ — Liubich-Kupzov,
5) Kolmogorov, and
6) Stein: $p = q = r = 1, T = \mathbb{R}$.

The article [4] written by the author in collaboration with Magaril-Il'yaev is devoted to consideration of these cases and their generalizations. Here we give some comments to results of [4].

In [4] the approach we spoke about is applied to the following generalization of the problem of Hardy-Littlewod-Polya and Taikov. Let D^α, $\alpha \in \mathbb{R}^d$ be an operator of α-th derivarive in $\mathbb{R}^d$ in the sence of H. Weyl ($D^\alpha x(\cdot) = F^{-1}E^\alpha F x(\cdot)$, where F is Fourier transform, F^{-1} is inverse mapping and E^α is the operator of multiplication $E^\alpha \tau = \prod_{j=1}^{d} (i\tau_j)^{\alpha_j}$). Consider the problem

$$(**) \qquad \|D^{\alpha^0} x(\cdot)\|_{L_p(\mathbb{R}^d)} \to \max, \quad (p = 2, \infty), \quad \|D^{\alpha^j} x(\cdot)\|_{L_2(\mathbb{R}^d)} \leq \gamma_j.$$

Let us consider at the beginning the case $p = 2$.

1) *Generalized Hardy-Littlewood-Polya problem.* Denote $(2\pi)^d |Fx(t)|^2 dt = d\mu(t)$. Then the problem $(**)$ with $p = 2$ have the following formalization:

$$(1) \qquad -\int_{\mathbb{R}^d} |\tau|^{2\alpha^0} d\mu(\tau) \to \min, \quad \int_{\mathbb{R}^d} |\tau|^{2\alpha^j} d\mu(\tau) \leq \gamma_j^2, \ \ \mu \geq 0.$$

It is the problem of linear progravmming. Application of the duality method immediately leads to solution of the problem.

The same method solves the analogous problems on other manifolds, for example, on $\mathbf{S}^d$.

2) *Generalized Taikov's problem.* After Fourier transform we obtain the following formalization of the problem (∗∗) with $p = \infty$:

$$(2) \qquad -\int_{\mathbb{R}^d} Fx(\tau)dt \to \min, \int_{\mathbb{R}^d} |\tau|^{2\alpha^j} |Fx(\tau)|^2 dt \le \gamma^2/(2\pi)^d.$$

This is the problem of quadratic programming. Duality method reduces (2) to a finite dimensional problem, and if $N = d+1$ the answer could be expressd in an explicit form.

3) *Gabusin's case.* Here is one of the possible formalizations of Gabushin's problem:

$$(3) \qquad x^{(k)}(0) \to \min, \quad \int_{\mathbb{R}_+} (x^2 + (x^{(n)})^2)dt \le 1.$$

It is a convex problem of calculus of variations. After applying Lagrange principle we obtain a linear differential equation of the $2n$-th order and transversality conditions, which give the possibility to find (the unique) solution of the equation. From convexity of the problem this solution gives absolute minimum of (3).

4) *Liubich-Kupzov's case.* One of the formalizations of the problem is similar to the previous:

$$(4) \qquad \int_{\mathbb{R}_+} (x^{(k)}(t))^2 dt \to \max, \quad \int_{\mathbb{R}_+} (x^2 + (x^{(n)})^2)dt \le 1.$$

The Lagrange function of the problem has the form

$$L(x(\cdot)) = \int_{\mathbb{R}_+} ((x^{(n)})^2 - \lambda(x^{(k)})^2 + x^2)dt,$$

A solution of Euler equation $(-1)^n x^{(2n)} - \lambda(-1)^k x^{(k)} + x = 0$. together with transversality conditions leads to a solution which depends on λ. Fhe following identity takes place (k_j are roots of characteristic equation $(-1)^n z^{2n} - \lambda(-1)^k z^{2k} + 1 = 0$, lying at the left halfspace):

$$\int_{\mathbb{R}_+} (x^{(n)} + \sum_{j=1}^{n-1} k_j(\lambda) x^{(n-j)} + x)^2 dt + Q_\lambda(x(0), \dots, x^{(n-1)}(0)),$$

where Q_λ is a quadratic form. (This identity could be checked directly, although it is nothing else but the general formula of Weierstrass in calculus of variations.) The form Q_λ is positive when $\lambda = 0$ and not nonnegative if λ is large enough. The solution corresponds to $\widehat{\lambda}$ for which Q_λ is nonneganive.

5) *Kolmogorov's case.* It is the most substantial case. Using the method we are speaking about it is possible to solve many new Kolmogorov-type problems connected with extrapolation of smooth functions, inequalities for derivatives of smooth functions at a fixed point, recovery of smooth functions etc.

Stein's case is unique when this method does not solve the problem. (Stein reduced the problem to Kolmogorov's case).

The method we discussed has an extremely wide circle of applications. But limits of this ideology is bounded by limits of Lagrange principle. The most important class of problems where the form of Lagrange principle is unknown and must be modified is the class of multidimentional (where $t \in \mathbb{R}^d$) versions of problems being considered in this paper (on extrapolation, inequality of derivatives at fixed points, Kolvogorov's type inequalities in d-dimensional case and so on).

REFERENCES

[1] A. D. Ioffe, V. M. Tikhomirov. Theory of Extremal Problems. North-Holland,1979.

[2] V. M. Tikhomirov. Harmonics and splines as optimal tools of approximation and recovery. *Uspehi Mat. nauk* **50**, *6* (1995).

[3] V. B. Demidovich, G. G. Magaril-Il'yaev, V. M. Tikhomirov. Extrema of linear functionals on Tchebishev subspases. In preparation.

[4] G. G. Magaril-Il'yaev, V. M. Tikhomirov. On Kolmogorov-type inequalities for derivatives. *Matematicheskii sbornik* **189**, *1* (1998).

Faculty of Computing Mathematics & Cybernetics
Moscow State University
119 899 Moscow
Russia
e-mail: `tikh@tikhomir.mccme.ru`